

CORRECTING TYPE IA SUPERNOVA DISTANCES FOR SELECTION BIASES AND
CONTAMINATION IN PHOTOMETRICALLY IDENTIFIED SAMPLESR. KESSLER^{1,2}, D. SCOLNIC^{1,3}*draft in prep. October 18, 2016*

ABSTRACT

We present a new technique to create a bin-averaged Hubble Diagram (HD) from photometrically identified SN Ia data. The resulting HD is corrected for selection biases and contamination from core collapse (CC) SNe, and can be used to infer cosmological parameters. This method, called “Bias Corrected Distances” (BCD), includes two fitting stages. The first BCD fitting stage combines a Bayesian likelihood with a Monte Carlo simulation to bias-correct the fitted SALT-II parameters, and also incorporates CC probabilities determined from a machine learning technique. The BCD fit determines 1) a bin-averaged HD (average distance vs. redshift), and 2) the nuisance parameters α and β , which multiply the stretch and color (respectively) to standardize the SN brightness. In the second stage, the bin-averaged HD is fit to a cosmological model where priors can be imposed. We perform high precision tests of the BCD method by simulating large (150,000 event) data samples corresponding to the Dark Energy Survey Supernova Program (DES-SN). Our tests include three models of intrinsic scatter, each with two different CC rates. In the BCD fit, the SALT-II nuisance parameters α and β are recovered to within 1% of their true values. In the cosmology fit, we determine the dark energy equation of state parameter w using a fixed value of Ω_M as a prior: averaging over all six tests based on $6 \times 150,000 = 900,000$ SNe, there is a small w -bias of 0.006 ± 0.002 . There is an additional w -uncertainty due to the assumed cosmology in the simulations; after iterating, this uncertainty is roughly equal to $\sigma_w/7$ where σ_w is the uncertainty of the data. For spectroscopically confirmed SN Ia samples, we compare the BCD method to other HD fitting methods, and we explain a long-standing paradox in which highly biased results are obtained if the Gaussian normalization term, $-2\ln(\sigma_\mu)$, is included in the HD chi-squared likelihood: the solution is that bias corrections are needed to obtain unbiased results. Finally, the BCD fitting code is publicly available in the SNANA package.

Subject headings: techniques: cosmology, supernovae

1. INTRODUCTION

The discovery of the accelerating expansion of the universe (Riess et al. 1998; Perlmutter et al. 1999) using Type Ia supernovae (SN Ia) has motivated increasingly large transient searches in broadband imaging surveys. Approximately 1,000 spectroscopically confirmed SNe Ia (Conley et al. 2011; Betoule et al. 2014; Scolnic et al. 2014a; Rest et al. 2014) have been combined with measurements of the cosmic microwave background (Planck Collaboration XVI 2014; Hinshaw et al. 2013) to measure the dark energy equation of state parameter (w) and today’s matter density (Ω_M). There are not enough spectroscopic resources to dramatically increase the sample size of *confirmed* SNe Ia, and therefore several programs are aiming to acquire very large samples of photometrically identified SNe Ia to measure dark energy properties with increased precision. These SN programs include the recently completed Pan-STARRS1 (Kaiser et al. 2002), the ongoing Dark Energy Survey Supernova Program (DES-SN: Bernstein et al. 2012), and the Large Synoptic Survey Telescope (LSST: Ivezić et al. 2008; LSST Science Collaboration et al. 2009), expected

to begin in the next decade.

These photometric samples are expected to include contamination from core collapse (CC) SNe, and this contamination must be accounted for in the inference of cosmological parameters. Biases from selection effects and light curve fitting must also be accounted for. In this paper we present a new HD fitting method, called “bias-corrected distances” (BCD), for extracting bias-corrected cosmological parameters from a photometrically identified SN Ia sample. We test BCD on high quality simulations of the DES-SN program. The SALT-II framework (Guy et al. 2010) is used for light curve fitting, and a nearest neighbor (NN) method is used for photometric classification. We assume that an accurate spectroscopic redshift is obtained from the host galaxy, and ignore the small fraction of wrong SN-host matches described in Gupta et al. (2016). BCD is not sensitive to the particular method of photometric classification, but is sensitive to how well the resulting contamination is modeled by simulations. Although BCD is designed to treat photometrically identified samples, it is also applicable to spectroscopically confirmed samples by simply leaving out the CC term in the likelihood.

In addition to this effort to improve HD fitting, there are several other Bayesian fitting techniques under development which use simulations to correct for selection effects (Weyant et al. 2013; Rubin et al. 2015; Shariff et al. 2016). While we do not make an explicit

kessler@kicp.uchicago.edu

¹Kavli Institute for Cosmological Physics, University of Chicago, Chicago, IL 60637, USA²Department of Astronomy and Astrophysics, University of Chicago, 5640 South Ellis Avenue, Chicago, IL 60637, USA³Hubble, KICP Fellow

recommendations for the best method, the validation of any method should include high-statistics tests with realistic simulated SN light curves that are fit with the SALT-II model.

Recent SN Ia cosmology analyses are based on a light curve fit for each event using the SALT-II model, which determines the best-fit values for the overall amplitude (x_0), stretch (x_1), and color (c). The stretch and color describe rest-frame properties of the SN Ia that are needed to standardize the brightness, and the amplitude describes the observed SN brightness and dimming from the distance modulus.

The ensemble of fitted $\{x_0, x_1, c\}$ values are passed to a second stage of Hubble-diagram (HD) fitting. The HD fit simultaneously determines the cosmological parameters (e.g., w and Ω_M) and standardization coefficients, α and β , which multiply the stretch and color, respectively, in order to standardize the brightness and determine a distance for each event. Marriner et al. (2011, hereafter M11) introduced another HD fitting method to determine α and β without simultaneously fitting for the cosmological parameters, and to compute a cosmology-independent distance modulus for each event. This “SALT2mu” method fixes the cosmological parameters and fits for a distance modulus offset ($\Delta_{\mu,z}$) in multiple redshift bins. The extraction of cosmological parameters can be obtained by fitting the distance moduli vs. redshift, or fitting $\Delta_{\mu,z}$ vs. redshift. The advantage of the SALT2mu output is that a wide variety of cosmology models and priors can be employed in subsequent analyses without repeatedly fitting for the nuisance parameters (α, β).

HD chi-squared fitting based on fitted SALT-II parameters is fundamentally flawed because it does not account for biases from selection effects, light curve fitting, and CC contamination. With spectroscopically confirmed SN Ia samples, the effect from biases has been evaluated in recent analyses (Conley et al. 2011; Betoule et al. 2014; Scolnic et al. 2014a) by running the same flawed procedure on both data and a simulation, and using the simulated sample to measure the average distance modulus bias in redshift bins. This bias-vs-redshift was applied as a correction to the data in a second-iteration HD fit.

This iterative correction is conceptually flawed for three reasons. First, Scolnic & Kessler (2016, hereafter SK16) have shown that correcting biases as a function of redshift is not a full description, and that the proper correction requires a 3-dimensional function of redshift, stretch, and color. Second, the standardization parameters α and β were determined without bias corrections, and then used to simulate bias corrections vs. redshift. The third issue is more of an implementation flaw rather than a conceptual flaw: simulations in previous analyses were generated with approximate stretch and color populations that were not rigorously determined as in SK16.

The basic concept of BCD is to analyze the fitted SALT-II parameters (x_0, x_1, c) by maximizing a likelihood function that combines elements from the 1) BEAMS methodology from Kunz et al. (2007) and Hlozek et al. (2012, hereafter H12), to form a likelihood from the combined Ia and CC probabilities, 2) SALT2mu program (M11), to fit for a distance modulus offset in redshift

bins, and 3) SNANA simulations (Kessler et al. 2009), to determine biases. The simulation is also used to determine the shape of the CC probability map as a function of distance modulus and redshift, which should be more accurate than the analytical approximation used in H12.

One of the original goals of the SALT2mu program is to convert the fitted SALT-II light curve parameters into a distance modulus for each SN Ia, so that the resulting HD can be fit with arbitrary cosmology-fitting programs. This strategy does not work with CC contamination because cosmology-fitting programs implicitly assume that all events are genuine SNe Ia. Using the BCD likelihood, however, the fitted distance modulus offset in each redshift (z) bin is properly corrected for CC contamination. The resulting $\Delta_{\mu,z}$ -vs- z function can be fit with arbitrary cosmology models, maintaining the original spirit of the SALT2mu program.

As part of developing BCD, we address a long-standing paradox about the HD-fit χ^2 formalism. Within the SALT-II framework, the uncertainty on the distance modulus (σ_μ) depends on fitted nuisance parameters (α, β) and thus a Gaussian normalization term, “ $-2\log(\sigma_\mu)$,” should be added to the χ^2 function that is minimized. It has been long recognized, however, that adding the Gaussian normalization term results in large biases on the fitted parameters (e.g., see Appendix B of Conley et al. (2011)). For a spectroscopically confirmed SN Ia sample, we verify that adding the Gaussian normalization term does indeed result in large biases if bias corrections are ignored. Including both the Gaussian normalization term along with bias corrections results in good parameter estimates.

The analysis presented here includes only statistical uncertainties, in order to check the precision of the BCD method. We therefore assume that our simulation correctly predicts CC contamination, measurement noise, SN Ia intrinsic scatter, and selection biases. An analysis of real data, however, must characterize the accuracy of the simulations and include sources of inaccuracy in the calculation of systematic uncertainties.

An overview of the paper is as follows. We begin with a review of the SALT-II framework in §2. Simulations are described in §3. The photometric analysis and nearest neighbor method are described in §4. The BCD method is presented in §5, with result in §6 and §7. A comparisons with other HD fitting methods is given in §8, and we conclude in §9. All simulation and analysis codes used in this paper are publicly available in the SNANA package,⁴ including analysis input files.⁵

2. REVIEW OF SALT-II FRAMEWORK

Here we briefly review key aspects of the SALT-II and SALT2mu framework, both of which are critical components of BCD. As described in §1, the fitted light curve parameters for each SN are epoch of peak brightness (t_0), amplitude (x_0), stretch (x_1), and color ($c \simeq B - V$ at t_0). For each event the fitted parameters are used to standardize the SN brightness and determine a distance modulus using the Tripp relation (Tripp 1998),

$$\mu = m_B + \alpha x_1 - \beta c - M_0 \quad (1)$$

⁴ <http://snana.uchicago.edu>

⁵ \$SNANA_ROOT/sample_input_files/KS2016

where $m_B = -2.5 \log(x_0)$, M_0 is the rest-frame magnitude for an SN Ia with $x_1 = c = 0$, and α, β are global nuisance parameters to standardize the SN Ia brightness.

After the light curve fits, the next step is a global fit comparing each measured distance modulus (μ) to a model distance that depends on redshift and cosmological parameters, $\mu_{\text{model}} = -2.5 \log(d_L/10\text{pc})$, where for a flat w CDM universe ($\Omega_M + \Omega_\Lambda = 1$),

$$d_L(z, w, \Omega_M) = (1+z) \frac{c}{H_0} \int_0^z \frac{dz'}{E(z')}, \quad (2)$$

$$E(z) = \left[\Omega_M(1+z)^3 + \Omega_\Lambda(1+z)^{3(1+w)} \right]^{1/2}.$$

The **SALT2mu** program uses **MINUIT**⁶ to minimize the HD chi-squared function

$$\chi_{\text{HD}}^2 = \sum_i (\mu_i - \mu_{\text{model},i} - \Delta_{\mu,z})^2 / \sigma_{\mu,i}^2 \quad (3)$$

$$\sigma_\mu^2 = \sigma_{\text{int}}^2 + (\sigma_\mu^z)^2$$

$$+ C_{m_B, m_B} + \alpha^2 C_{x_1, x_1} + \beta^2 C_{c, c}$$

$$+ 2\alpha C_{m_B, x_1} - 2\beta C_{m_B, c} - 2\alpha\beta C_{x_1, c},$$

$$\sigma_\mu^z = \left(\frac{5}{\ln(10)} \right) \frac{1+z}{z(1+z/2)} \sqrt{\sigma_z^2 + (\sigma_{\text{pec}}/c)^2}.$$

The fitted parameters are $\alpha, \beta, \sigma_{\text{int}}$ and a distance offset ($\Delta_{\mu,z}$) in each redshift bin. In the definition of the uncertainty term, σ_μ , the event index i has been dropped. C is the fitted covariance matrix among the $\{m_B, x_1, c\}$ parameters, σ_{int} is the intrinsic scatter term, σ_z is the redshift uncertainty, and σ_{pec} is the peculiar velocity uncertainty. The **SALT2mu** program can add an arbitrary 3×3 intrinsic scatter matrix (Σ) in the calculation of σ_μ ; here we use $\Sigma_{m_B, m_B} = \sigma_{\text{int}}^2$ and set all other Σ terms to zero. This choice can be interpreted as including only the coherent scatter model (COH model in §3) in the distance uncertainties, and ignoring intrinsic variations in color and stretch.

Rather than fitting for cosmological parameters (w, Ω_M) appearing in the $\mu_{\text{model},i}$ term, the cosmology parameters are fixed to reference values ($w_{\text{ref}}, \Omega_{M,\text{ref}}$) and the $\Delta_{\mu,z}$ are fit in N_z redshift bins. For typical fits the redshift bin size is < 0.1 . For all event indices i whose redshift lies in the same bin, $\Delta_{\mu,z}$ is fixed to the same value. The key assumption here is that within each redshift bin, the local shape of the HD is well described by the reference cosmological model, and that the difference is characterized by an offset ($\Delta_{\mu,z}$). Thus, instead of fitting for two cosmology parameters (w, Ω_M), **SALT2mu** returns N_z fitted $\Delta_{\mu,z}$ values along with α, β , and σ_{int} . M_0 in Eq. 1 is defined as the average of the $\Delta_{\mu,z}$ values, and each $\Delta_{\mu,z} \rightarrow \Delta_{\mu,z} - M_0$.

Note that the χ_{HD}^2 in Eq. 3 has no Gaussian normalization term, $-2 \ln(\sigma_\mu)$. Minimizing this quantity will be referred to as the “traditional χ_{HD}^2 ” method.

After obtaining the fitted $\Delta_{\mu,z}$ from the **SALT2mu** program, cosmological parameters are obtained in a separate fit of $\mu_{\text{ref},z} + \Delta_{\mu,z}$ vs. redshift, where $\mu_{\text{ref},z}$ are the distances computed from the reference cosmological parameters ($w_{\text{ref}}, \Omega_{M,\text{ref}}$) used in the **SALT2mu** fit. The **SALT2mu**

program could in principle report an average distance modulus ($\mu_z \equiv \mu_{\text{ref},z} + \Delta_{\mu,z}$) in each redshift bin, with no impact on subsequent cosmology fitting codes. However, we prefer to report the fitted $\Delta_{\mu,z}$ instead.

3. SIMULATIONS

We test BCD on **SNANA** simulations based on the cadence and observing conditions from the first DES-SN season. The SN survey component of DES is described in Kessler et al. (2015), and the Dark Energy Camera (DECam) is described in Flaugher et al. (2015). The survey consists of ten 3 deg^2 fields, observed roughly once per week in each *griz* passband. Eight of these “shallow” fields are observed to an average depth of 23.5; the other two “deep” fields are observed to an average depth of 24.5. The first DES-SN season was used to build a library consisting of sky noise, zero point, and point-spread function (PSF) for each observation in the ten fields. This library is used to simulate realistic light curves at random times and sky locations. For each observation, the simulated magnitude is converted into a flux using the image zero point and CCD gain. The simulated flux uncertainty is computed from the PSF, sky noise, and zero point.

For SNe Ia, the redshift-dependent volumetric rate (R) is taken from Dilday et al. (2008), with $R(z) \propto (1+z)^{1.5}$. For better statistical constraints on the w -precision of the BCD method, an artificial low-redshift ($z < 0.08$) sample is added, which comprises $\sim 10\%$ of the total sample as shown in Fig. 1. This low- z sample is generated with the same *griz* passbands and depth as for the DES-SN sample, and is therefore an ideal anchor with minimal selection bias.

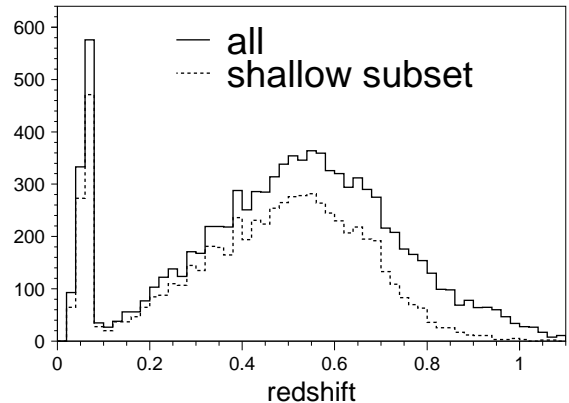


FIG. 1.— Simulated redshift distribution after selection requirements described in §4. The low-redshift ($z < 0.1$) subset is $\sim 10\%$ of the total.

SN Ia model magnitudes are generated from the **SALT-II** light curve model. To match the observed Hubble residual dispersion, three different models of intrinsic scatter are studied: 1) “COH” model of coherent, or achromatic scatter, with $\sigma_{\text{int}} = 0.13 \text{ mag}$, 2) “G10” scatter from the **SALT-II** model in which 70% of the contribution to the Hubble residuals comes from achromatic variation and 30% comes from chromatic variation, and 3) “C11” model from Chotard et al. (2011) in

⁶ www.cern.ch/minuit

which 25% of the contribution to the Hubble residuals comes from achromatic variation and 75% comes from chromatic variation.

The COH model is not consistent with observations, but is included as a test because the HD fitting model with constant σ_{int} is equivalent to the COH model, and thus one might naively expect that simulating the COH model would yield less biased results. The latter two models are broadband-variation models based on observations, and were converted into spectral variation models in Kessler et al. (2013) so that intrinsic variation is simulated by varying the underlying SALT-II spectra. These models add intrinsic scatter without changing the average underlying SN Ia model; the average generated SN Ia flux after intrinsic variations are applied is the same as the SALT-II model flux with no intrinsic scatter. The simulated values of the standardization parameters (Table 1) are $\alpha_{\text{SIM}} = 0.14$ and $\beta_{\text{SIM}} = 3.2$ for the COH and G10 scatter models; for the C11 model $\beta_{\text{SIM}} = 3.85$.

The underlying color and stretch populations are each described by an asymmetric Gaussian distribution defined by three parameters. For the color distribution,

$$P(c) \equiv \exp[-(c - \bar{c})^2 / 2\sigma_-^2] \quad (c \leq \bar{c}) \quad (4)$$

$$P(c) \equiv \exp[-(c - \bar{c})^2 / 2\sigma_+^2] \quad (c > \bar{c})$$

where \bar{c} is the value with maximum probability, and σ_- and σ_+ are the low- and high-sided Gaussian widths. A similar parameterization describes the stretch (x_1) distribution. We use the parameters in the High- z row in Table 1 of SK16, which are shown here in Table 1 for each model of intrinsic scatter. Note that in a more realistic analysis, the redshift dependent populations should be used.

TABLE 1
SIMULATION PARAMETERS FOR SN Ia PROPERTIES

Model	α_{SIM}	β_{SIM}	Color Params: \bar{c} σ_- σ_+			Stretch Params: \bar{x}_1 σ_- σ_+		
COH ^a	0.14	3.20	-0.054	0.043	0.101	0.973	1.472	0.222
G10	0.14	3.20	-0.054	0.043	0.101	0.973	1.472	0.222
C11	0.14	3.85	-0.099	0.003	0.119	0.964	1.467	0.235

^aSK16 did not evaluate population parameters for the COH model, so here we use the G10 parameters.

CC SN types II and Ibc are generated using the volumetric rate from Bazin et al. (2009), with a redshift dependence of $(1+z)^{3.6}$. Simulated light curves are generated from spectra that have been mangled to match photometric observations of 42 CC light curves as described in Kessler et al. (2010, hereafter K10). The relative fraction and luminosity function (LF) for each subtype (II, Ib/c) are from Li et al. (2011, hereafter L11). In K10, the simulation selects a random template spectrum and applies a magnitude offset and random Gaussian scatter such that the LF matches L11. The brightness distribution of the original CC light curves is preserved, and the LF from L11 is achieved with an additional random magnitude scatter. Here we use the same technique for the Type-II SNe, but alter the procedure for Type Ib to account for one anomalously bright event which over estimates the contribution from bright events. The

brightness for each Ib template spectrum is adjusted to have the mean LF brightness, and the full LF spread is simulated with random scatter (Jones et al., 2016, in prep).

Since a spectroscopic host galaxy redshift is required in this analysis, we use the spectroscopic matching efficiency estimated in Bernstein et al. (2012).⁷ This efficiency drops to 0.75 at $z = 0.5$, and drops to 0.50 at $z = 0.95$.

The following effects are available options in the SNANA simulation, but have been left out for this study: peculiar velocities, weak lensing, host-galaxy correlations, and redshift-dependent SN properties (e.g., α , β , population parameters).

Because the simulation generates realistic light curve fluxes and uncertainties, there are no assumptions about the analytical form of detection thresholds, and it properly simulates arbitrarily complex survey-selection triggers based on the number of detections, as well as the distribution of detections over nights and passbands. It also accounts for time-dependent variations due to weather and instrumental effects. We fit each simulated light curve with the same SALT-II model (and code) used on data, and thus light curve fitting biases are included such as those found in SK16.

4. NEAREST NEIGHBOR METHOD FOR PHOTOMETRIC CLASSIFICATION

Here we describe the analysis to photometrically select SNe Ia. The overall strategy is to first apply selection cut-windows, or “box cuts”, on several analysis variables to reduce the CC/Ia fraction to $< 10\%$. The second step is to apply the nearest neighbor method (Sako et al. 2014, hereafter referred to as NN) to further reduce the CC contamination and to determine a Type Ia probability needed for the BCD likelihood (§5). Numerous machine learning (ML) methods can be applied to photometric classification as described in Lochner et al. (2016). Our choice of NN is arbitrary and adequate to test BCD; we make no claim about which ML method is best.

After fitting the light curves with the SALT-II model, the box cuts are as follows:

1. at least one observation with $T_{\text{rest}} < -2$ days, where T_{rest} is the rest-frame epoch with respect to the epoch of peak brightness.
2. at least one observation with $T_{\text{rest}} > +10$ days.
3. at least three bands have an observation with signal-to-noise ratio (S/N) above 5.
4. redshift $z < 1.10$.
5. $|x_1| < 3$ and $|c - 0.1| < 0.4$.
6. the SALT-II light curve fit probability (P_{fit}), computed from the fit- χ^2 and number of degrees of freedom, satisfies $P_{\text{fit}} > 0.05$.

The first four requirements ensure good light curve quality, while the last two requirements select SNIa-like light curves. At this stage of the analysis we transition to

⁷ See Table 18 column with $m < 24$ and $\kappa_{\text{Ia}} = 0.5$.

the more sophisticated NN method to further reduce CC background.

Our NN analysis is based on the 3-dimensional space of $\{z, x_1, c\}$, where z is a spectroscopic host galaxy redshift and the latter two variables are from the SALT-II light curve fit. For a given data event,⁸ the basic idea is to define a sphere centered at $\{z, x_1, c\}$, generate a large simulated sample, and count the number of Type Ia and CC SNe that land inside the sphere. The event is classified to be the SN type representing the majority inside the sphere. A procedure called ‘NN training’ determines the optimal size of the sphere based on maximizing the product of the efficiency and purity.

More formally, for each event in the data sample, the NNs are events from a large simulated training sample that satisfy a 3-dimensional distance constraint,

$$d^2 = \left[\frac{(z - z')^2}{D_z^2} + \frac{(c - c')^2}{D_c^2} + \frac{(x_1 - x'_1)^2}{D_{x1}^2} \right] < 1, \quad (5)$$

where the primed quantities are from the simulated training sample. The optimal distance-metric parameters (D_z, D_c, D_{x1}) are determined from a training procedure which maximizes the product of the SN Ia purity and the efficiency. The final selection requirement is that for simulated neighbors satisfying Eq. 5, more than half are true SNe Ia with at least 1σ confidence. Thus if there are 20 neighbors, at least 13 ($\sigma = 2.1$) must be true SNe Ia to be classified as SN Ia. If only 12 ($\sigma = 2.2$) of the neighbors are true SNe Ia, the 1σ -above-half requirement fails and the event is classified as unknown. The Type Ia probability for each event, $P_{\text{NN,Ia}}$, is defined as the fraction of NN training events (satisfying Eq. 5) that are true SNe Ia.

After passing the box cuts and NN requirement, there are a total of 10,000 events in the simulated data sample, and 70,000 events in each of the two training samples.⁹ With our nominal estimate of the CC rate, the CC contamination¹⁰ is 0.054 after the selection requirements, and drops to 0.015 after the NN requirement. The corresponding SN Ia loss from the NN requirement is 0.003. Since the CC contamination in a photometrically identified sample is not yet known, we test the BCD method with two different CC rates. Simulations are generated with our best estimate of the CC rate ($R_{\text{CC}} = 1$), and again with $3\times$ the CC rate ($R_{\text{CC}} = 3$). R_{CC} is defined here as the simulated CC rate divided by our best estimate of the rate. For each R_{CC} value, the CC contamination before and after the NN requirement is shown in Table 2. Fig. 2 shows the distance modulus residuals to illustrate the contamination before and after the NN requirement.

A final caveat is that the NN training has been performed on the combined deep+shallow fields, but in principle a separate training on deep and shallow sub-samples could be more optimal. However, since the same requirements are applied to the data and simulated samples, the level of NN optimization has no impact on the BCD per-

formance.

TABLE 2
CC CONTAMINATION FRACTION AND SN Ia EFFICIENCY VS. R_{CC} , BEFORE AND AFTER NN REQUIREMENT

R_{CC}	CC/All no NN	CC/All with NN	Eff(SNIa) with NN
1	0.054	0.015	0.997
3	0.146	0.035	0.992

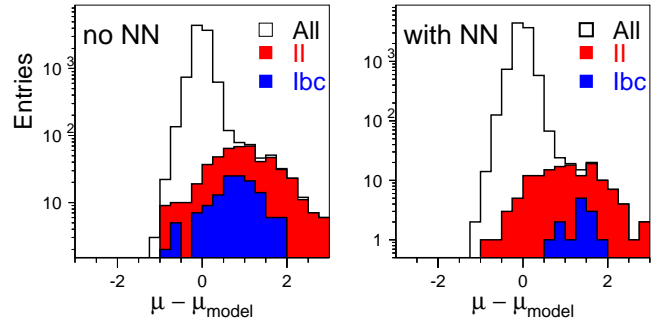


FIG. 2.— Distance modulus residual ($\mu - \mu_{\text{model}}$) for DES-SN simulation with $R_{\text{CC}} = 1$: after box cuts (left) and after NN requirement (right). Legend indicates contribution from all SN types (Ia+CC), and from types II and Ibc.

5. BCD METHOD

The BCD method consists of two fitting stages. The first step is the BCD fit to determine $\Delta_{\mu,z}$ in each redshift bin, where $\Delta_{\mu,z}$ is the distance-modulus offset with respect to an arbitrary cosmological model, which will be referred to as the reference cosmology. In addition to requiring light curve fit results (x_0, x_1, c) from the data, the BCD fit also requires input from a detailed simulation based on a rigorous characterization of the survey. The second step is to fit $\Delta_{\mu,z}$ -vs-redshift in order to determine the cosmological parameters. This step requires no understanding of the survey, and hence can be performed with a variety of fitting codes and priors.

5.1. Overview

Using the BEAMS formalism, we form a likelihood function consisting of Type Ia and CC probabilities. The joint likelihood for N events is $\mathcal{L} \equiv \prod_{i=1}^N \mathcal{L}_i$, where i is the event index, and MINUIT is used to minimize $-2\ln(\mathcal{L})$. The BCD likelihood for each event is

$$\mathcal{L}_i \equiv \left[\frac{P_{\text{NN,Ia}}^i}{P_{\text{tot}}^i} D_{\text{Ia}}(z_i, \mu_i) \right] + \left[\frac{S_{\text{CC}}^i (1 - P_{\text{NN,Ia}}^i)}{P_{\text{tot}}^i} D_{\text{CC}}(z_i, \mu_i) \right], \quad (6)$$

where z_i is the redshift and μ_i is the distance modulus. Dropping the event index i ,

$$D_{\text{Ia}} = \text{probability function} : \exp[-\chi_{\text{HD}}^2/2]/\sigma_{\mu}\sqrt{2\pi} \quad (7)$$

$$D_{\text{CC}} = \text{probability map from simulation.} \quad (8)$$

$$P_{\text{tot}} = P_{\text{NN,Ia}} + S_{\text{CC}}(1 - P_{\text{NN,Ia}}) \quad (9)$$

⁸ We use the term ‘data’ here, even though it is a simulated data sample, since the NN procedure is the same with real data.

⁹ One of the training samples serves as an independent data sample to avoid statistical anomalies from self-training.

¹⁰ CC contamination is defined as the fraction of events which are true CC.

D_{Ia} and D_{CC} are the normalized Ia and CC distributions such that at any redshift the distance modulus integrals satisfy

$$\int_{-\infty}^{+\infty} D_{\text{Ia}}(z, u) d\mu = \int_{-\infty}^{+\infty} D_{\text{CC}}(z, u) d\mu = 1, \quad (10)$$

and $\int \mathcal{L}_i d\mu = 1$ for each event. D_{Ia} (Eq. 7) depends on the traditional χ_{HD}^2 term defined in Eq. 3, and includes the fitted $\Delta_{\mu, z}$. $P_{\text{NN, Ia}}$ is the NN probability for Type Ia, $(1 - P_{\text{NN, Ia}})$ is the NN probability for CC, and S_{CC} is a fit parameter allowing for an arbitrary scale of the CC probability. Note that $S_{\text{CC}} \neq 1$ implies that $P_{\text{tot}} \neq 1$, but the overall normalization of the likelihood, $\int \mathcal{L}_i d\mu = 1$, is always satisfied. Although we use a specific case of NN probabilities in the likelihood, this method works using probabilities from any classification method.

The fitted parameters in \mathcal{L} are α , β , σ_{int} , S_{CC} , and $\Delta_{\mu, z}$. Compared to minimizing the traditional χ_{HD}^2 , the only additional fit parameter in \mathcal{L} is S_{CC} .

We make two improvements with respect to the BEAMS analysis in H12: 1) the SALT-II fitted parameters (m_B, x_1, c) are bias-corrected based on a simulation of SNe Ia, and 2) the analytical functional form for D_{CC} in H12 is replaced by a simulated probability map to remove assumptions about the analytical form. The subsections below describe these improvements in detail.

While the traditional χ_{HD}^2 (Eq. 3) has no Gaussian normalization term to avoid biases, D_{Ia} (Eq. 7) in the BCD likelihood must include this term to properly normalize the Ia and CC probabilities. Thus the Gaussian normalization term cannot be arbitrarily removed from the BCD likelihood, as has been done with the traditional χ_{HD}^2 method. The correct solution is to include bias corrections as described in §5.2; this issue is examined further in §8.1.

We refer to the simulated data sample as ‘data’, to clearly distinguish the data from a large independent simulated sample used for bias corrections and the CC likelihood. Thus the descriptions are valid when replacing a simulated data sample with real data.

Finally, since the shallow and deep-field samples differ in depth by 1 mag, bias corrections and the CC likelihood are determined separately for each sub-sample. This separation of the corrections illustrates the more general principle of combining multiple samples from different surveys.

5.2. Bias Corrected Distance

For each data event, a bias-corrected distance is defined by replacing the fitted parameters (m_B, x_1, c) in Eq. 1 with bias-corrected parameters,

$$\begin{aligned} \mu^* &= m_B^* + \alpha x_1^* - \beta c^* - M_0 \\ &= (m_B - \bar{\delta}_{m_B}) + \alpha(x_1 - \bar{\delta}_{x_1}) - \beta(c - \bar{\delta}_c) - M_0 \end{aligned} \quad (11)$$

where the star superscript indicates a bias-corrected quantity. The bias corrections $(\bar{\delta}_{m_B}, \bar{\delta}_{x_1}, \bar{\delta}_c)$ are determined from a large ‘BiasCor’ simulation with 500,000 events after the requirements in §4. For any given event, we cannot exactly determine the true parameter bias (δ) because of variations caused by intrinsic scatter and measurement noise. We therefore interpolate the bias ($\bar{\delta}$) in a 5-dimensional (5D) space of $\{z, x_1, c, \alpha, \beta\}$. The cell sizes

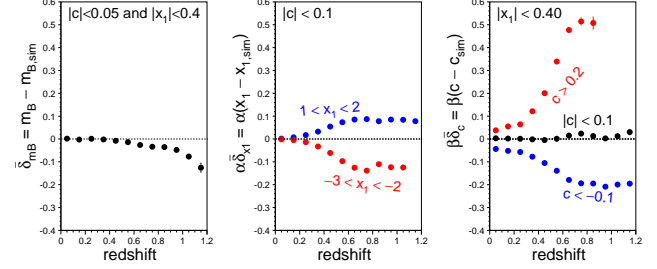


FIG. 3.— Bias corrections $\bar{\delta}_{m_B}$, $\alpha\bar{\delta}_{x_1}$, and $\beta\bar{\delta}_c$ are shown as a function of redshift. The pre-factors α, β are used to show the bias in distance-modulus magnitudes. The parameter selection ranges are shown on each panel.

are $\{0.05, 0.50, 0.05\}$ for $\{z, x_1, c\}$, respectively. The bias correction has a very weak dependence on α and β , and therefore a 2×2 grid is adequate with two fixed α values ($\alpha_{\text{SIM}} \pm 0.04$) and two fixed β values ($\beta_{\text{SIM}} \pm 0.4$), where $\alpha_{\text{SIM}}, \beta_{\text{SIM}}$ are given in Table 1. The SN Ia parameters α and β are generated at discrete values because there is no evidence for a distribution. Each 3D sub-cell of $\{z, x_1, c\}$, however, includes a continuous distribution based on the SN rate vs. redshift and the population of stretch and color.

Each $\bar{\delta}$ -bias value ($\bar{\delta}_{m_B}, \bar{\delta}_{x_1}, \bar{\delta}_c$) is determined by linear interpolation in the 5D space. The bias and grid-location in each 5D cell is the weighted-average among all BiasCor events in the cell, and each weight is σ_μ^{-2} . A valid BiasCor cell requires at least 3 events, and at least 3 valid cells are required for interpolation. Data events with fewer than 3 cells are rejected, which reduces the CC contamination as described in §7.

There is a subtle interpolation issue for events in the highest redshift bin ($1.05 < z < 1.10$). Recall that the bias value in each redshift bin is defined at the weighted average grid-location, \bar{z}_{bin} , and thus for $\bar{z}_{\text{bin}} < z < 1.10$ the bias cannot be interpolated if there are no events beyond $z > 1.10$. While it is possible to extrapolate the bias for $z > \bar{z}_{\text{bin}}$, extrapolating a rapidly varying bias function could result in a large error. To avoid extrapolating, the BiasCor simulation includes events with $1.10 < z < 1.15$ to provide a reliable interpolation for events with $\bar{z}_{\text{bin}} < z < 1.10$.

The bias corrections as a function of redshift are illustrated in Fig. 3 for a few arbitrary stretch and color bins; the largest corrections are a few tenths of a mag for very blue and red colors. These corrections are similar to the 3D (z, x_1, c) corrections in SK16, except that here we account for the dependence on α and β as illustrated in Fig. 4.

5.3. Bias Corrected Distance Uncertainty

After applying bias corrections, the true Hubble residual scatter can be significantly reduced, and it can differ from the calculated uncertainty (σ_μ) in Eq. 3. This effect is shown in Fig. 5a,b for the deep and shallow field sub-samples, respectively. In the deep fields (Fig. 5a), bias corrections result in a $\sim 20\%$ reduction in the Hubble scatter at the highest redshifts, while in the shallow fields (Fig. 5b) the reduction reaches $\sim 40\%$. The calculated uncertainty (σ_μ) is in good agreement with the bias-corrected Hubble scatter in the deep sample, but σ_μ is much too large in the shallow sample.

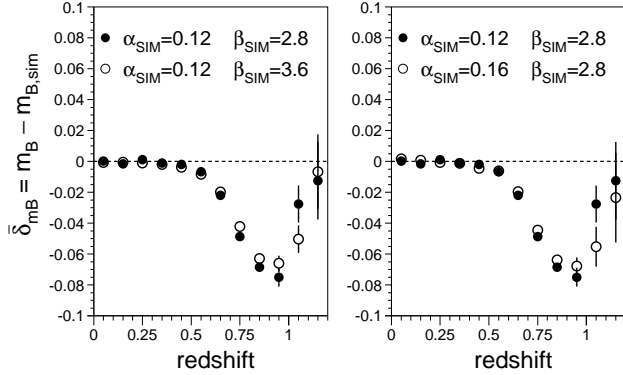


FIG. 4.— Bias correction $\bar{\delta}_{m_B}$ vs. redshift, for entire shallow field subset. Solid circles are the same in both panels. Open circles show bias with same α_{SIM} and different β_{SIM} (left), and with same β_{SIM} and different α_{SIM} (right). Note that the vertical scale is $\times 5$ smaller compared to Fig. 3.

To ensure a robust estimate of the distance uncertainty in the likelihood, we apply a 2 dimensional σ_μ correction as a function of redshift and color. We found little σ_μ dependence on the stretch (x_1) parameter, and thus did not include an x_1 correction. The color dependence is illustrated in Fig. 5c. The redshift binning is the same as for the bias corrections, but there are only 3 color bins in order to maintain better bin statistics for measuring the rms. The form of the correction is

$$\sigma_\mu^{\text{data}} \rightarrow \sigma_\mu^{\text{data}} \times R_\sigma \quad (12)$$

$$R_\sigma \equiv \text{rms}(\text{with biasCor})/\sigma_\mu^{\text{sim}}, \quad (13)$$

where R_σ is determined from the same BiasCor simulation used to determine the bias corrections, and is the ratio of blue and black curves in Fig. 5a,b. The simulated intrinsic scatter term (σ_{int}) is determined from the subset with signal-to-noise ratio $S/N > 60$, and it is important to verify that $R_\sigma \simeq 1$ at low redshifts with high S/N . $R_\sigma(z)$ is determined at each α - β grid point, as well as for deep and shallow sub-samples. During the minimization, R_σ is interpolated as a function of α, β, z, c .

An unexpected artifact of the bias correction is that the Hubble scatter at high redshift is somewhat smaller in the shallow field sub-sample than in the deep fields, which contradicts our naive expectation that the scatter should be smaller in the deep fields (compare blue-dashed curves in Figs. 5a,b). The reason for this paradox is that at high redshifts a narrower range of brightness is observed in the shallow fields, resulting in smaller scatter after bias corrections. We do not, however, get better measurements in the shallow fields for two reasons. First, there are more events per square degree and redshift bin in the deep fields. Second, the shallow field bias corrections are more sensitive to the cosmological parameters assumed in the BiasCor simulation, resulting in a larger systematic uncertainty (§6.1).

5.4. CC Probability Distribution

In H12, D_{CC} in Eq. 6 has the same functional form as D_{Ia} , except that $\mu_{\text{model}} \rightarrow \mu_{\text{model}} + \Upsilon(z)$ where $\Upsilon(z)$ is a polynomial function of redshift with coefficients as additional fit parameters. The error term (σ_μ) has the same functional form as in the SN Ia likelihood, but has an independent σ_{int} term. The Gaussian form of D_{CC} trivially

satisfies the normalization condition (Eq. 10), but may not be a sufficiently accurate model. Here we replace the analytical Gaussian form with a simulated map shown in the lower panels of Fig. 6. The normalization constraint in Eq. 10 is imposed numerically. The CC map in Fig. 6 clearly has discontinuities in the derivatives, which can cause problems with MINUIT minimization. To alleviate such issues, the mean and rms in each redshift bin are used to define a Gaussian function.

5.5. Determination of σ_{int}

Using the traditional χ_{HD}^2 method (Eq. 3), σ_{int} is defined such that $\chi_{\text{HD}}^2 = N_{\text{dof}}$, where N_{dof} is the number of degrees of freedom in the fit. However, with the Ia+CC likelihood in Eq. 6, there is no physical interpretation to defining σ_{int} such that $-2\log(\mathcal{L}) = N_{\text{dof}}$. To maintain the traditional interpretation of σ_{int} , we impose a constraint that $\chi_{\text{HD}}^2 = N_{\text{dof}}^{\text{Ia}}$ for the subset of “ $N_{\text{dof}}^{\text{Ia}}$ ” events that have a low CC probability:

$$P_{CC/Ia} \equiv \frac{S_{CC}(1 - P_{NN,Ia})D_{CC}(z, \mu)}{P_{NN,Ia}D_{Ia}(z, \mu)} < 0.2. \quad (14)$$

Ideally the σ_{int} subset would require $P_{CC/Ia} \ll 1$, but a more restrictive subset removes legitimate SNe Ia from the μ -residual tails and results in an underestimate of σ_{int} . In the SALT2mu program σ_{int} is evaluated iteratively rather than as a fit constraint. Convergence of $\chi_{\text{HD}}^2/N_{\text{dof}}^{\text{Ia}} = 1$ to within 1% typically requires 2-3 fit iterations.

5.6. Fit for Cosmological Parameters

The BCD method consists of two fitting stages. The first step is to determine $\Delta_{\mu,z}$ by minimizing the BCD likelihood in Eq. 6. The second step, determining the cosmological parameters, is described here. For a spectroscopically confirmed SN Ia sample, one can perform the cosmology fit with either the distances for each SN, or the $\Delta_{\mu,z}$. However, for a photometrically identified sample with CC contamination, the $\Delta_{\mu,z}$ are properly corrected while the individual distances are not.

As described in §2, the result of the BCD fit is a set of distance offsets vs. redshift bin, $\Delta_{\mu,z}$, using a reference set of cosmological parameters ($w_{\text{ref}}, \Omega_{M,\text{ref}}$) that are fixed in the BCD fit. The final fitted values ($w_{\text{fit}}, \Omega_{M,\text{fit}}$) are determined by minimizing

$$\chi_\Delta^2 = \sum_z (\Delta_{\mu,z} + \mu_{\text{ref},z} - \mu_{\text{model},z})^2 / \sigma_{\Delta\mu,z}^2. \quad (15)$$

$\Delta_{\mu,z}$ and $\sigma_{\Delta\mu,z}$ are the BCD-fitted distance offset and uncertainty in each redshift bin (z), $\mu_{\text{ref},z}$ is the distance modulus computed from the reference parameters w_{ref} and $\Omega_{M,\text{ref}}$, and $\mu_{\text{model},z}$ is the distance modulus computed from the floated parameters w and Ω_M . Within each redshift bin, $\mu_{\text{ref},z}$ and $\mu_{\text{model},z}$ are computed at the weighted-average redshift, where each weight is σ_μ^{-2} .

χ_Δ^2 is very similar to χ_{HD}^2 in Eq. 3, except that the sum over individual events is replaced with a sum over bin-averaged distances in redshift bins. The SALT2mu program, which implements the BCD fit and produces the $\Delta_{\mu,z}$, can also be used to minimize Eq. 15 with a suitable change in fitting options. However, it is advantageous to use more specialized cosmology fitting programs that include priors from other constraints or more

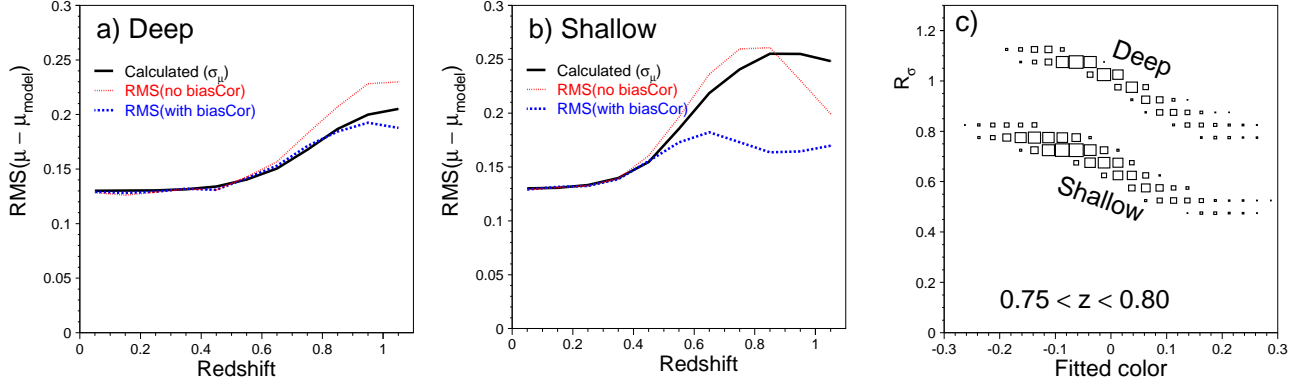


FIG. 5.— RMS of $\mu - \mu_{\text{model}}$ vs. redshift before bias corrections are applied (red-dotted), after bias corrections are applied (blue-dashed), and calculated σ_μ from Eq. 3 (black line). Left panel (a) is for the 2 deep fields, and middle panel (b) is for the 8 shallow fields. Right panel (c) shows the R_σ correction (Eq. 13) on σ_μ vs. fitted color, in the redshift bin indicated on the panel.

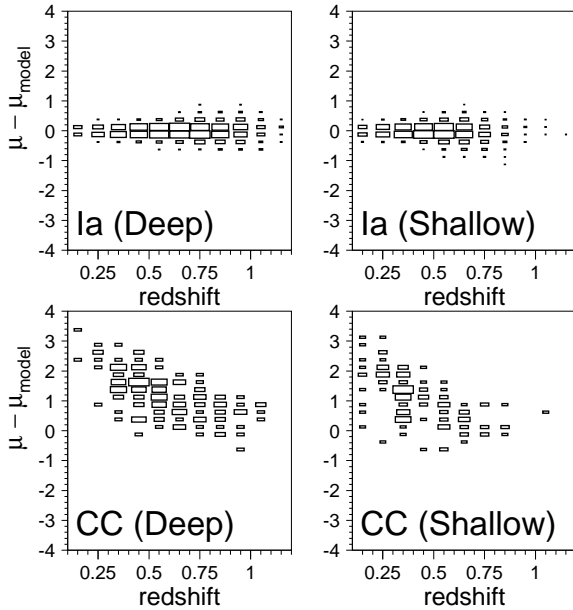


FIG. 6.— Distribution of distance residual vs. redshift, generated from the simulation for SN Ia (top) and CC SNe (bottom). Deep & shallow sub-samples are shown separately in the left and right panels. Box cuts plus the NN requirement have been applied. For the BCD likelihood, the simulated CC distribution in each redshift bin is normalized and replaced with a Gaussian. The SNIa distribution is shown for comparison, but the analytic SNIa likelihood is used instead.

diverse cosmological models. To focus on w -biases we impose a strong Gaussian prior on the matter density, $\Omega_M = \Omega_{M,\text{ref}} \pm 0.0001$, and a flat prior on w ($-1.5 < w < -0.5$).

6. RESULTS-I: TESTS WITH SIMPLIFIED SIMULATIONS

We begin with a simplified test of the BCD method such that any observed bias in the cosmology or nuisance parameters would point to a problem with the method. The simplifications in the simulated data and BiasCor samples are 1) exclude CC SNe, 2) generate coherent intrinsic scatter model (COH model in §3) so that the σ_{int} term in the BCD likelihood exactly matches the generated model of intrinsic scatter, and 3) remove Galactic reddening and peculiar velocities. In the light curve fitting, the SALT-

II model uncertainties are set to zero since there is no color variation in the intrinsic scatter model. Since there are no CC events, NN training is skipped and $S_{\text{CC}} = 0$ in the BCD likelihood (Eq. 6). In spite of these simplifications, the selection biases are still present and must be accounted for to obtain unbiased results.

To reach a sensitivity to a BCD-induced w -bias below 0.01, we simulate 15 independent data samples, each with $\sim 10,000$ events satisfying the box cuts and NN requirements in §4. The BiasCor sample has $\sim 500,000$ events, and is used to correct all 15 data samples. Distance moduli in all simulations are computed from a cosmological model with a flat universe and $\Omega_{M,\text{ref}} = 0.3$. We set $w_{\text{ref}} = -1$ for the NN training (§4), BiasCor sample, and BCD likelihood. For the data samples we set the true w -value, $w_{\text{data}} = -1$, and also test with other values to check the BCD method when the $w_{\text{ref}} \neq w_{\text{data}}$. The w -bias is defined as

$$w\text{-bias} \equiv w_{\text{fit}} - w_{\text{data}}, \quad (16)$$

where w_{fit} is described in §5.6. Fig. 7 shows the fitted $\Delta_{\mu,z}$, averaged over the 15 samples, vs. redshift. In the left panel, $w_{\text{data}} = -1$ and the fitted $\Delta_{\mu,z}$ agree well with the expected curve (dashed line through zero). In the next two panels of Fig. 7, $w_{\text{data}} = -0.9$ and -0.8 , which differs from $w_{\text{ref}} = -1$ in the BCD fit. The resulting $\Delta_{\mu,z}$ show a significant redshift dependence, and is expected as shown by the dashed curve. At higher redshifts, however, there is a notable discrepancy between the fitted $\Delta_{\mu,z}$ and the prediction, and this discrepancy is due to an incorrect bias correction induced by a slightly incorrect cosmology model. The maximum $\Delta_{\mu,z}$ discrepancy is roughly 0.01 mag per 0.1 difference between w_{data} and w_{ref} . The resulting w -bias is quantified below in §6.1, and a potential solution is discussed in §9.

For $w_{\text{data}} = -1$, the BCD-fitted parameters for the 15 samples are illustrated in Fig. 8 using an ideogram.¹¹ The parameters α , β , and σ_{int} are shown as ratios with respect to the true input values, with the weighted-average printed at the top of each plot. The α bias is almost 2%, while β and σ_{int} are recovered to within $\sim 1\%$. The w -bias is -0.002 ± 0.004 (lower-right panel of Fig. 8), and $\chi^2_{\text{red}} \simeq 0.5$ suggests that the fitted w -uncertainties

¹¹ <http://pdg.lbl.gov/2015/reviews/rpp2015-rev-rpp-intro.pdf>

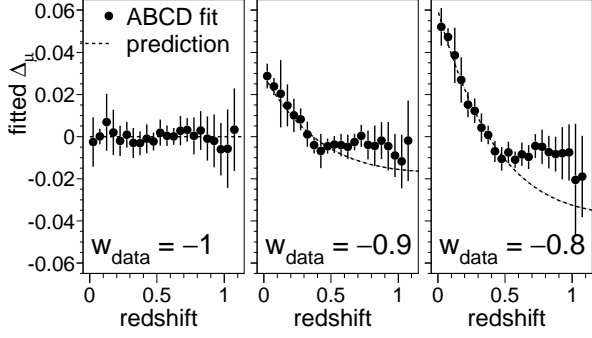


FIG. 7.— Average BCD-fitted value of $\Delta_{\mu,z}$ vs. redshift, and error bars show the rms spread among the 15 samples. The simulated w_{data} is indicated on each panel: $-1, -0.9, -0.8$ in the three panels, respectively. The same BiasCor sample, with $w_{\text{ref}} = -1$, is used in all three BCD fits. The dashed curve in each panel shows the prediction based on the Λ CDM model. The discrepancies between the BCD fit and prediction (two right panels) is due to the difference between w_{data} and w_{ref} .

may be overestimated, although the probability for such a low χ^2_{red} with correct uncertainties is 6.5%.

6.1. Cosmology-Dependent Bias in BiasCor Simulation

Here we illustrate a subtle bias from using incorrect cosmology parameters in the BiasCor simulation. For this test, w_{ref} is fixed to -1 in the BiasCor simulation, while w_{data} takes on different values in the data samples. For the three w_{data} values shown in Fig. 7, the BCD fit results and w -bias are shown in Table 3. The α bias persists at the 1-2% level, with large variations in χ^2_{red} (0.4-2). The β parameter is measured to within 1%, with a much smaller spread in χ^2_{red} values.

Table 3 shows a w -bias induced by the BiasCor dependence on cosmological parameters, and this bias increases non-linearly as $|w_{\text{data}} - w_{\text{ref}}|$ increases. Using the first two rows of Table 3 to estimate the local derivative, the w -bias is approximately given by $(w_{\text{ref}} - w_{\text{data}})/7$. This bias can be reduced with an iterative procedure in which w_{ref} is updated with the previous w_{fit} value, but the proximity of w_{ref} to the true value is limited by the total statistical+systematic uncertainty, σ_w . Since w_{ref} has the same uncertainty as w_{fit} , there is an additional and irreducible w -uncertainty of $\sim \sigma_w/7$ induced by the BiasCor dependence on cosmological parameters. The corresponding uncertainties in the deep and shallow field sub-samples are $\sigma_w/12$ and $\sigma_w/5$, respectively, and illustrates that this w -uncertainty depends on the SN sample.

To gain further insight we repeated the BCD method with $R_\sigma = 1$ (Eq. 13). This test corresponds to using the black curves in Figs. 5a,b, instead of the blue-dashed curves. We find that the bias is reduced by almost a factor of 2. With $R_\sigma = 1$ the distance uncertainties are larger at higher redshift, while the resulting σ_{int} is 7% smaller in order to satisfy the χ^2_{HD} constraint in §5.5. This change in uncertainties assigns greater weight to the BCD likelihood at lower redshifts and is thus less sensitive to the BiasCor simulation at higher redshift.

7. RESULTS-II: MORE REALISTIC TESTS

Here we test the BCD method on realistic simulations that include CC contamination, NN analysis, SN Ia intrinsic scatter models with color variation (G10 and C11 in §3), and Galactic reddening. In addition to the flux

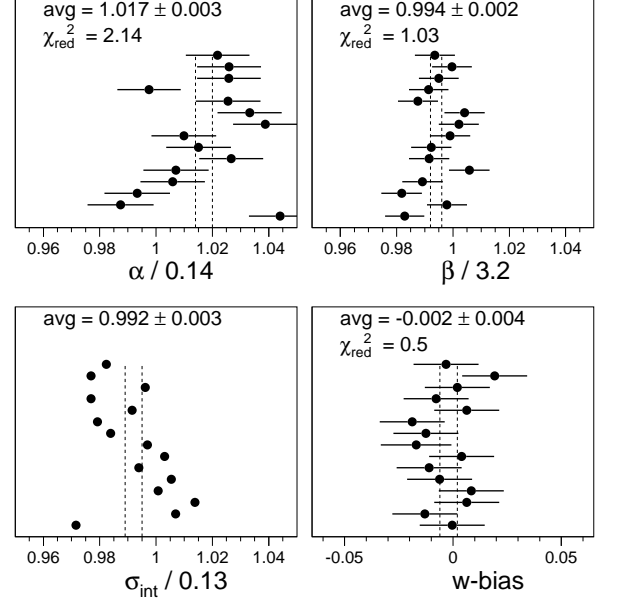


FIG. 8.— Results of 15 simulated data samples fit with the BCD likelihood, and $w_{\text{data}} = w_{\text{ref}} = -1$. First three plots show the ratio of the best-fit nuisance parameter ($\alpha, \beta, \sigma_{\text{int}}$) to the true generated value. Last plot shows the bias on w_{fit} ($w+1$). The “avg” on each panel shows the weighted average and weighted uncertainty, except for the σ_{int} panel where the uncertainty is $\text{rms}/\sqrt{15}$. χ^2_{red} on each panel is the reduced χ^2 determined from the 15 values, their uncertainties, and the weighted average. Vertical dashed lines show $\text{avg} \pm \text{weighted uncertainty}$.

uncertainties, all light curve fits include the SALT-II model uncertainties that are based on the G10 model. For the samples generated with the COH and C11 scatter models, the incorrect G10-based model uncertainties are still used in the light curve fits.

For each of the three intrinsic scatter models, 15 independent data samples are generated, each with 10,000 events passing selection requirements. The averaged results are thus based on 150,000 events for each intrinsic scatter model. A BiasCor sample with 500,000 events is generated for each intrinsic scatter model, and in the BCD fits we use the BiasCor sample with the correct intrinsic scatter model. Assuming the correct intrinsic scatter model is an optimistic assumption since we cannot clearly distinguish among the G10 and C11 models, but we leave these systematic tests to future analyses.

The data samples are generated with $w_{\text{data}} = -1$ ($= w_{\text{ref}}$), which is very close to the assumption that w_{ref} has been set to the first-iteration w_{fit} value: a caveat is that we have not included a statistical variation on w_{fit} .

The results are shown in Table 4. The first three rows show results based on the nominal CC rate ($R_{\text{CC}} = 1$) in the data and NN-training samples, and the resulting contamination is $\sim 1\%$. The next three rows show results with $R_{\text{CC}} = 3$, resulting in 2.5% contamination. Compared to $R_{\text{CC}} = 1$, the contamination with $R_{\text{CC}} = 3$ is less than $3\times$ higher because the NN training sacrifices a slightly higher SNIa loss for a lower contamination.

The BCD-fitted α and β are recovered to within 1%, and the w -bias is constrained at the level of 0.01. The CC probability scale factors, S_{CC} , are consistent with unity as expected. Interestingly, the w -bias seems smaller with the larger CC rate, albeit with low significance. The w -bias averaged over all six entries in Table 4, which

TABLE 3
BCD FIT RESULTS AND w -BIAS AVERAGED OVER 15 IDEAL DES-SNIA SAMPLES WITHOUT CC CONTAMINATION ^{a,b}

w_{data}	$\alpha/0.14$	χ^2_{red}	$\beta/3.2$	χ^2_{red}	$\sigma_{\text{int}}/0.13$	$w\text{-bias}^{\text{d}}$	χ^2_{red}
-1 ^c	1.017(3)	2.1	0.994(2)	1	0.992(3)	-0.002(4)	0.5
-0.9	1.016(3)	0.4	1.000(2)	0.9	1.000(3)	-0.017(4)	0.4
-0.8	1.020(3)	1.4	1.006(2)	0.7	1.014(3)	-0.044(4)	0.8

^a Each value is the weighted average and value in () is the weighted uncertainty in the last digit.

^b χ^2_{red} is the reduced χ^2 for the average among the 15 samples.

^c Results for $w_{\text{data}} = -1$ are also shown in Fig. 8.

^d $w_{\text{ref}} = -1$ in each BiasCor sample; w -bias can be reduced by iterating with $w_{\text{ref}} \rightarrow w_{\text{fit}}$.

TABLE 4
BCD FIT RESULTS AND w -BIAS AVERAGED OVER 15 REALISTIC DES SN(Ia+CC) SAMPLES^{a,b}

Intrinsic scatter model	R_{CC}	True CC fraction ^c	$\alpha/\alpha_{\text{SIM}}$	χ^2_{red}	β/β_{SIM}	χ^2_{red}	σ_{int}	S_{CC}	χ^2_{red}	$w\text{-bias}$	χ^2_{red}
COH	1	0.009	1.010(3)	1.0	0.997(2)	1.3	0.108	1.03(4)	1.8	0.015(4)	1.3
G10	1	0.010	1.002(4)	1.3	0.999(2)	1.1	0.075	1.02(4)	2.0	0.011(4)	1.0
C11	1	0.010	1.013(3)	1.3	0.997(2)	1.4	0.103	1.09(4)	1.4	0.000(3)	1.1
COH	3	0.023	1.001(3)	0.9	1.003(2)	2.7	0.103	1.00(2)	1.0	0.003(4)	1.6
G10	3	0.025	1.008(3)	0.5	1.005(2)	1.9	0.069	1.02(2)	0.9	0.002(3)	0.7
C11	3	0.026	1.016(3)	0.9	1.007(2)	2.1	0.096	1.04(2)	1.2	0.006(3)	0.8

^a Same as in Table 3

^b Same as in Table 3

^c True $N_{\text{CC}}/(N_{\text{CC}} + N_{\text{Ia}})$ after selection requirements.

includes 900,000 simulated SNe, is 0.006 ± 0.002 .

To get a sense of the importance of the CC term in the BCD likelihood (Eq. 6), we have run the BCD fits without the CC term by forcing $S_{\text{CC}} = 0$. For the data sample generated with $R_{\text{CC}} = 1$, the fitted α and β are biased by about 30% and the w -bias is 0.1 to 0.15 depending on the intrinsic scatter model. For the data sample with $R_{\text{CC}} = 3$, the w -bias is around 0.2. It is therefore essential to include an accurate CC term in the BCD likelihood.

As an additional test on the impact of CC contamination, we have removed the CC subset ($R_{\text{CC}} = 0$), corresponding to a spectroscopically confirmed (Ia-only) sample, and fit with only the SN Ia term in the BCD likelihood. For the COH and G10 scatter models, the Ia-only w -bias values are within 0.001 of those obtained from the photometric sample with CC contamination. For the C11 scatter model, the w -bias values differ by 0.004.

A subtle issue is how the BCD fit further reduces the CC contamination. In §4 (Table 2) we showed that the NN requirement significantly reduces the CC contamination: from 0.054 to 0.015 for $R_{\text{CC}} = 1$, and from 0.146 to 0.035 for $R_{\text{CC}} = 3$. However, the actual CC contamination values after the BCD fit (Table 4) are about 50% smaller: 0.01 and 0.025, respectively, with a corresponding SN Ia loss of 2%. After the NN requirement, the BCD fit further reduces the CC contamination because the bias correction imposes an implicit selection requirement that is similar to the NN requirement, but more strict. Recall that the NN requirement rejects data events where more than half the simulated neighbors are true CC SNe, and neighbor events are counted within a 3D sphere defined by $\{z, x_1, c\}$. The bias correction rejects events for which there are not enough SN Ia events in a rectangular $\{z, x_1, c\}$ -cell to determine the correction. The BiasCor

cell volume is $\sim 4\times$ smaller than the 3D sphere volume determined from the NN training, and thus requiring a valid bias correction is more strict.

We further check the effectiveness of the BiasCor requirement in reducing CC contamination by repeating the BCD analysis without the NN requirement, but still using the NN probabilities ($P_{\text{NN,Ia}}$ in Eq. 6). With $R_{\text{CC}} = 1$, the CC contamination fractions are about 10% higher compared to the combined NN+BiasCor requirement. With $R_{\text{CC}} = 3$, the CC fractions are 20% higher. For both R_{CC} values, the w -bias values are nearly identical with and without the NN requirement. Using only the NN requirement results in 50% more CC contamination, and therefore the BiasCor requirement is somewhat more effective than the NN requirement at reducing CC contamination, but at a cost of 2% additional loss of SN Ia. This result should not be interpreted as a general statement about the inferiority of the NN method, because the NN method can also be used to obtain lower CC contamination in exchange for reduced SN Ia efficiency. For example, replacing the “purity \times efficiency” metric with “purity ^{p} \times efficiency” ($p > 1$) reduces the CC contamination.

8. COMPARISON WITH OTHER BIAS CORRECTION METHODS

Here we compare the BCD method with other HD-fitting methods. First, in §8.1 we show the impact from using incorrect fitting methods: leaving out the Gaussian normalization and/or the distance-bias corrections. Next, in §8.2 we discuss previous analyses which used a redshift dependent distance-bias correction.

8.1. Incorrect Fitting Methods

Several recent SN Ia analyses are based on the traditional χ^2_{HD} approach (Eq. 3) to HD fitting (Conley et al. 2011; Betoule et al. 2014; Scolnic et al. 2014a; Rest et al. 2014). While some approaches have included a redshift-dependent bias correction (e.g., §8.2), the Gaussian normalization term, $-2\ln(\sigma)$, has never been included because it results in large biases (e.g., see Appendix B of Conley et al. (2011)). The underlying problem is that both the Gaussian normalization term and bias corrections are needed, and leaving out either results in a large bias as shown in Table 5. The first two rows of Table 5 show the effect of including the Gaussian normalization term without bias corrections: the fitted α and β are $\sim 20\%$ below their true values, and the w -bias is greater than $+0.1$. The next two rows show the effect of including bias corrections without the Gaussian normalization term: the fitted α and β are $\sim 30\%$ above their true values, and the w -bias is almost -0.1 . The last two rows show the effect of leaving out the Gaussian normalization and bias corrections: the fitted α and β are 5-10% away from their true values and the w -bias is ~ 0.04 . It is interesting to note that leaving out either the Gaussian normalization or bias corrections results in large fit biases with opposite signs; leaving out both corrections results in some cancellation and significantly smaller biases. Using the BCD method results in much smaller biases as shown in Table 3.

TABLE 5
HD FIT RESULTS FROM INCORRECT FITTING METHODS^a

scatter model	$\alpha/\alpha_{\text{SIM}}$	β/β_{SIM}	σ_{int}	w -bias
GN ^b =yes and BiasCor=no				
G10	0.787(3)	0.797(2)	0.104	0.118(4)
C11	0.786(3)	0.653(1)	0.120	0.124(4)
GN ^b =no and BiasCor=yes				
G10	1.281(4)	1.268(2)	0.059	-0.088(4)
C11	0.931(4)	1.332(3)	0.084	-0.085(4)
GN ^b =no and BiasCor=no				
G10	1.068(3)	0.958(2)	0.091	0.037(4)
C11	1.082(4)	0.781(2)	0.110	0.039(4)

^a Averaged over 15 simulated DES-SNIa samples without CC contamination.

^b GN = Gaussian normalization term in HD fit.

8.2. Redshift Dependent Bias Corrections

The most recent and precise SN Ia cosmology results (Betoule et al. 2014; Scolnic et al. 2014a) are based on the traditional χ^2_{HD} fitting approach with no Gaussian normalization term, but they accounted for distance biases using simulations. Recognizing that their HD fitting approach results in biased results, their strategy was to measure the average μ -bias (δ_μ) as a function of redshift by analyzing simulations with their biased HD fitting procedure that has no Gaussian normalization term and no bias corrections. Each distance modulus in the data is then corrected as a function of redshift, $\mu \rightarrow \mu - \delta_\mu(z)$, which we call a 1D BiasCor to distinguish from the 5D BiasCor in the BCD method. The final cosmology fit is applied to the 1D-corrected distances.

With an unrealistic assumption of using the correct α and β in the BiasCor simulation, the 1D BiasCor works as well as the 5D BCD method, but with 9% larger uncertainty on cosmological parameters. An important caveat is that previous analyses did not have a robust method for determining α and β , which are needed as inputs to the BiasCor simulation. In Betoule et al. (2014), α and β were determined from the traditional χ^2_{HD} fit, and thus may have biases as illustrated for DES in the last two rows of Table 5. Scolnic et al. (2014a,b) used a similar approach, and analyzed additional simulations to correct for the large β -bias in the C11 model.

To illustrate the impact of incorrect α and β , we generated a BiasCor simulation with $\alpha = 0.15$ and $\beta = 3.0$, which corresponds roughly to the biased values using the traditional χ^2_{HD} approach (last 2 rows in Table 5). We use this BiasCor simulation to apply a 1D correction to the simulated data samples from §6 that were generated with the true parameter values ($\alpha = 0.14$, $\beta = 3.2$): the resulting w -bias is 0.016 ± 0.004 , significantly larger than the 5D BCD result in the first row of Table 3.

The fundamental issue is not so much the 1D-vs-5D correction, but rather that previous analyses bias-corrected the distance ($\mu \rightarrow \mu - \delta_\mu$) instead of correcting each fitted parameter (m_B, x_1, c) shown in Eq. 11 for the BCD method. The δ_μ correction requires α and β to be pre-determined, while the BCD method explicitly updates δ_μ at each minimization step as α and β are varied.

It would be interesting to apply the BCD method to existing data samples in order to measure α and β , and to check for potential errors on the 1D distance-bias corrections and cosmological parameters. However, evaluating the α, β bias in previous analyses is beyond the scope of this work.

9. DISCUSSION AND CONCLUSION

The BCD fit has been implemented within the existing SALT2mu program; it is publicly available in SNANA, and ready to use on real data samples for which reliable simulated samples are available. This new program has been rigorously tested on large simulated data samples ($\sim 150,000$ events) with different models of intrinsic scatter and different CC rates. The CPU time for each BCD fit is less than 10 minutes for 10,000 events, and this time scales with the number of events. A BiasCor simulation sample of 500,000 events ($z < 1.2$) takes about 50 CPU-hours to create,¹² and a few hours are needed to generate a Ia+CC sample and perform the NN training. Below we describe some unresolved issues and areas for improvement.

The first unresolved issue is how to determine sub-samples for which the BiasCor simulation is run. The deep and shallow fields in DES-SN, with a 1 mag difference in search depth, is a rather obvious choice for separating the samples. However, for a survey with a fixed exposure time there are weather variations which change the search depth, and edge effects that truncate light curves; it is thus not clear if or how the BiasCor simulation should be split into multiple samples. Our current method for splitting samples is essentially trial-and-error. In this analysis, for example, we started with

¹² The BiasCor CPU breakdown is 10 hours for SN Ia generation, and almost 40 hours for light curve fitting.

shallow and deep field samples combined into a single BiasCor simulation, and found w -biases in the range of 0.01 to 0.02; generating separate BiasCor simulations for shallow and deep fields reduced the w -bias to below 0.01.

The second unresolved issue is that the BCD method does not currently incorporate systematic uncertainties other than the naive method of re-running the analysis for many systematic parameter variations. A more formal approach using covariance matrices (Conley et al. 2011) can in principle be added to the SALT2mu program. However, with sample sizes growing into the thousands, the covariance method could become computationally challenging. It is not yet clear if the best approach is to update the SALT2mu program to include systematic uncertainties, or to incorporate bias corrections in other cosmology-fitting methods that are better suited for including systematic uncertainties.

To incorporate systematic uncertainties, there is a potentially interesting alternative to BCD called “Approximate Bayesian Computation” (Weyant et al. 2013, ABC), where the basic idea is to find the set of cosmological parameters such that the resulting simulated sample best matches the data. To account for systematic uncertainties, the simulated samples include variations in systematics parameters such as calibration zero points, CC rate, SN Ia model parameters, etc. This method differs from BCD because ABC has no explicit assumption about the likelihood. It is similar to BCD in that both methods rely on accurate simulations to model observations. However, while the BCD method relies on a fixed set of simulations generated before fitting, ABC generates simulations for each set of parameters and thus does not suffer from the cosmology-dependent bias illustrated in Table 3 and Fig. 7.

Inferring cosmological parameters with a full treatment of systematic uncertainties is likely to require significant computing resources, and thus limit the number of analysis iterations. Since the BCD method is relatively fast, it is useful to develop methods for characterizing systematic uncertainties, and also to decompose the relative contribution from numerous sources of uncertainty.

Another issue is that uncertainties on the assumed cosmology in the BiasCor simulation introduce an additional w -uncertainty of $\sigma_w/7$ for the DES-SN sample (§6 and Fig. 7). This finding raises the prospect that the highest redshift events could add a BiasCor uncertainty which exceeds the reduced statistical uncertainty.

A more practical problem is the potential need to regenerate the BiasCor sample with updated cosmology results. This iterative procedure would render the $\Delta_{\mu,z}$ output usable only to those who can regenerate large BiasCor simulations, and thus negate the goal of enabling arbitrary groups to use arbitrary cosmology fitting codes on the $\Delta_{\mu,z}$ output. This goal can be partially preserved by generating BiasCor samples for a grid of cosmological parameters (\vec{C}) such as $\vec{C} = \{w, \Omega_M\}$ or $\vec{C} = \{w_0, w_a, \Omega_M\}$, where $w(a) = w_0 + w_a(1 - a)$ and “ a ” is the scale factor for the expanding universe. Generating a BiasCor grid with two values per cosmology parameter requires 2^2 and 2^3 BiasCor simulations for the two models, respectively, which only requires a few hundred CPU hours. Arbitrary cosmology fitting programs could interpolate the BCD-fitted $\Delta_{\mu,z}$ as a function of \vec{C} , with the

limitation of using only the cosmological model defined by \vec{C} . While a truly model-independent output from BCD is desired, it is not clear how to achieve this.

A number of enhancements to the BCD likelihood are feasible with minor code changes. Additional fitted parameters in the CC component would add flexibility, such as allowing the Gaussian mean and width in each redshift bin (Fig. 6) to be adjusted by a polynomial function of redshift. Fitted polynomial coefficients different from zero would indicate a discrepancy between the data and CC simulation. The SN Ia component of the likelihood could be enhanced with additional parameters to describe host-galaxy correlations, redshift dependent SN properties, and quadratic stretch and color terms in the Tripp equation. Lastly, the interpolated bias corrections are linear between the nearest 5D grid nodes where the biases are defined. Spline interpolation may work better, as long as the CPU time and memory are not significantly increased.

The population parameters describing the asymmetric color and stretch distributions have been determined in a separate analysis (SK16). However, it may be possible to include these parameters in the BCD fit by multiplying the likelihood by a simulated probability characterizing the number of events in stretch and color bins. The downside of this enhancement is that the SALT-II parameter biases cannot be determined before the fit, and thus could result in a significant increase in CPU time. So while a simultaneous fit of all parameters is an attractive goal, it may be more computationally efficient to iteratively evaluate the population parameters separately from the BCD fit.

In summary, we have presented a new SN Ia cosmology fitting method, BCD, which uses a large simulation to account for biases from sample selection, light curve fitting, and CC contamination. With an infinitely precise prior on Ω_M , the BCD method introduces a w -bias below 0.01. A proper cosmology analysis, however, should characterize and account for uncertainties on the BiasCor and CC simulations.

10. ACKNOWLEDGEMENTS

This work was supported in part by the Kavli Institute for Cosmological Physics at the University of Chicago through grant NSF PHY-1125897 and an endowment from the Kavli Foundation and its founder Fred Kavli. We gratefully acknowledge support from NASA grant 14-WPS14-0048. RK is supported by DOE grant DE-AC02-76CH03000. DS is supported by NASA through Hubble Fellowship grant HST-HF2-51383.001 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555.

REFERENCES

- Bazin, G., Palanque-Delabrouille, N., Rich, J., et al. 2009, *A&A*, 499, 653
- Bernstein, J. P., Kessler, R., Kuhlmann, S., et al. 2012, *ApJ*, 753, 152
- Betoule, M., Kessler, R., Guy, J., et al. 2014, *A&A*, 568, A22
- Chotard, N., Gangler, E., Aldering, G., et al. 2011, *A&A*, 529, L4
- Conley, A., Guy, J., Sullivan, et al. 2011, *ApJS*, 192, 1
- Dilday, B., Kessler, R., Frieman, J. A., et al. 2008, *ApJ*, 682, 262
- Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, *arXiv:1504.02900*
- Gupta, R. R., Kuhlmann, S., Kovacs, et al. 2016, *ArXiv e-prints*
- Guy, J., Sullivan, M., Conley, A., et al. 2010, *A&A*, 523, A7
- Hinshaw, G., Larson, D., Komatsu, et al. 2013, *ApJS*, 208, 19
- Hlozek, R., Kunz, M., Bassett, B., et al. 2012, *ApJ*, 752, 79
- Ivezic, Z., Tyson, J. A., Abel, B., et al. 2008, *arXiv:0805.2366*
- Kaiser, N., Aussel, H., Burke, B. E., et al. 2002, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Vol. 4836, *Survey and Other Telescope Technologies and Discoveries*, ed. J. A. Tyson & S. Wolff, 154–164
- Kessler, R., Bassett, B., Belov, P., et al. 2010, *PASP*, 122, 1415
- Kessler, R., Bernstein, J. P., Cinabro, D., et al. 2009, *PASP*, 121, 1028
- Kessler, R., Guy, J., Marriner, J., et al. 2013, *ApJ*, 764, 48
- Kessler, R., Marriner, J., Childress, M., et al. 2015, *AJ*, 150, 172
- Kunz, M., Bassett, B. A., & Hlozek, R. A. 2007, *Phys. Rev. D*, 75, 103508
- Li, W., Leaman, J., Chornock, R., et al. 2011, *MNRAS*, 412, 1441
- Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., & Winter, M. K. 2016, *ArXiv e-prints*
- LSST Science Collaboration, Abell, P. A., Allison, J., Anderson, S. F., Andrew, J. R., Angel, J. R. P., Armus, L., Arnett, D., Asztalos, S. J., Axelrod, T. S., & et al. 2009, *arXiv:0912.0201*
- Marriner, J., Bernstein, J. P., Kessler, R., et al. 2011, *ApJ*, 740, 72
- Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, *ApJ*, 517, 565
- Planck Collaboration XVI. 2014, *A&A*, 571, A16
- Rest, A., Scolnic, D., Foley, R. J., et al. 2014, *ApJ*, 795, 44
- Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, *AJ*, 116, 1009
- Rubin, D., Aldering, G., Barbary, K., Boone, K., Chappell, G., et al. 2015, *ApJ*, 813, 137
- Sako, M., Bassett, B., Becker, A. C., et al. 2014, *arXiv:1401.3317*
- Scolnic, D. & Kessler, R. 2016, *ArXiv e-prints*
- Scolnic, D., Rest, A., Riess, A., et al. 2014a, *ApJ*, 795, 45
- Scolnic, D. M., Riess, A. G., Foley, R. J., et al. 2014b, *ApJ*, 780, 37
- Shariff, H., Jiao, X., Trotta, R., & van Dyk, D. A. 2016, *ApJ*, 827, 1
- Tripp, R. 1998, *A&A*, 331, 815
- Weyant, A., Schafer, C., & Wood-Vasey, W. M. 2013, *ApJ*, 764, 116